Brainwave: A Multidisciplinary Journal (ISSN: 2582-659X), Vol. 3, Special Issue, March 2022, pp. 25-30, © Brainware University

Challenges and Application of Masked and Unmasked face recognition using FaceNet

Shiplu Das*#
sld.cs@brainwareuniversity.ac.in

Protim Pal[#]
bwubts18075@brainwareuniversity.ac.in

Jayanta Saha[#]
bwubts18086@brainwareuniversity.ac.in

Soham Ghosh[#]
bwubts18014@brainwareuniversity.ac.in

Suman Patra[#]
bwubts18012@brainwareuniversity.ac.in

[#] *Dpartment of CSE, Brainware University, Barasat, West Bengal, India*

Abstract:

The COVID-19 pandemic is an unprecedented crisis leading to a huge number of casualties. In order to reduce the spread of the Corona Virus and to protect themselves, people often wear masks. In this era the face recognition techniques are suffering due to they have been designed for no-mask face recognition. The need has aroused for an application which is capable of recognizing faces covered with face masks. In this paper, we proposed a reliable method based on IOT and Neural Network in order to address the problem of face recognition with masks. The very first step of this method will be to discard the masked face region. Then, in order to reduce the resolution reduction during the sub-sampling process, dilated convolution is used. In the last stage, the feature information from the image of the face is used in the attention mechanism neural network to reduce the loss of information in the sub-sampling process and to improve the rate of the face in the image being recognized correctly. When doing the experiments, we have used the RMFRD and SMFRD databases of Wuhan University to evaluate and compare the recognition rate. The results from the experiments suggest the proposed algorithm to have a better recognition rate.

Keywords

Computer Vision, Face Recognition, Masked Face Recognition, mask and no-mask face recognition, Deep Learning, Conv. Nets, Classification, KNN, COVID-19.

Introduction

In recent years, The COVID-19 virus can be spread through contact and contaminated surfaces, therefore, the classical biometric systems based on passwords or fingerprints are not anymore safe. So in this situation, Face recognition is safer than any other touchable device. We all know that corona virus has stopped spreading by wearing a face mask. So, when we wear the mask it has some problems. The problems are - Criminals take advantage of the mask and do illegal work which is very adverse to us. It is very hard to identify when a person is wearing a mask. When one person wears a mask then the face recognition methods are not used properly because the whole face image is not given. It is important that the opening of the nose area for the task of face recognition otherwise it is used for face normalization, pose correction, and face matching. Due to these problems, it is very hard to recognize faces using face recognition methods. To avoid these problems, we can use two different tasks namely: masked face detection and masked face recognition. For this part we have used pre-trained model [5] for good encoding of the images. For doing the discrimination task, we used the K-NN classifier on the encoding obtained from an intermediate layer of the pre-trained model.

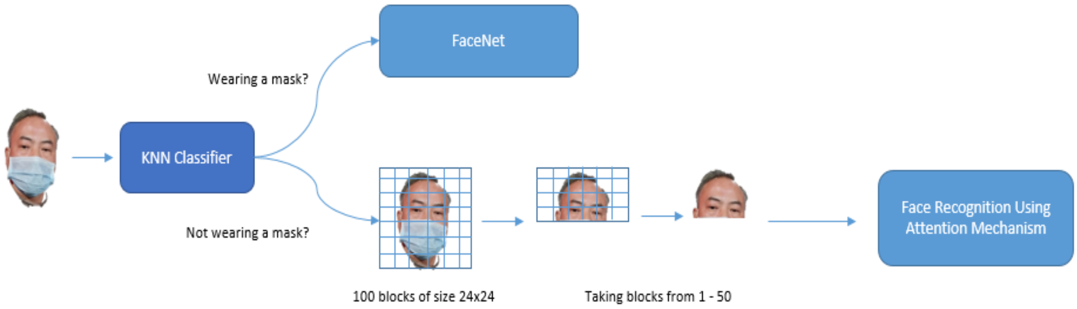

Fig 1: The proposed mechanism for mask and no-mask face recognition

Different challenges of Masked and Unmasked face recognition process:

Discrimination based on the presence of mask

First step in the system is to check whether the face in the image is wearing a mask or not. Conditioned on this factor the subsequent process is determined. We have used transfer learning for these purpose. As for a pre-trained model we have chosen MobileNet for its efficiency and ability to encode small details. To be specific, we have used the “conv_preds” intermediate layer to get a good encoding of the face image and added it with an appropriate label to the KNN classifier with K = 2.

Using FaceNet for non-masked faces

After the KNN classification stage if the image comes out to be a non-masked one then it is fed into an implementation of the FaceNet for getting a good encoding of the image. Since FaceNet encodings captures the features that can be used for finding similarities between two faces, while doing the face recognition task we can do 2 passes through the network and finally a dense layer to classification purpose. More clearly, in the first pass we can use the image that is registered in the system against a person and for the second image we can use the captured image. Then after getting the two encodings for two images we can use the dense layer to determine if those encodings are of the same person or not.

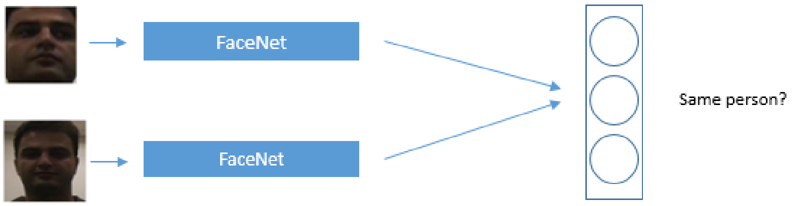

Fig. 2: Face Recognition using FaceNet

Attention Mechanism for Face Recognition

The overall network architecture used for the recognition of cropped face regions from the previous stage, instead of using the traditional U-shaped structure, it uses two paths for secondary sampling. For obtaining more detailed information Dilated convolution is. In order to extract features from feature context information, we have used ResNet. To imitate a similar mechanism as to the human eye, attention mechanism is implemented [6]. Finally, to integrate the collected features of different sensory fields to obtain better results, a feature fusion module is designed.

(1) Dilated Convolution: Dilated convolution in the spatial information path is used to maintaining a fixed field of view as well as to improve the resolution. In formal terms, ordinary convolution can be described as following:

$$P(x,y)\cdot G(x,y)=\sum_{d_1=0}^{\omega}\sum_{d_2=0}^{m} K(d_1,d_2)\cdot P(x-d_1,y-\ d_2) \tag{1}$$

In the equation, P (x, y) represents the pixel value at the point (x, y) of the original image and G (x, y) is notation for values of convolution kernel multiplied by it with the size of$\omega\ \times\ m$.

Whereas, the equation for the dilated convolution goes as follows:

$$P(x,y)\cdot G'(x,y)=\sum_{d_1=0}^{\omega}\sum_{d_2=0}^{m} G'(d_1,d_2)\cdot P(x-k\times d_1,y-\ k\times d_2) \tag{2}$$

Where, k and $G'(x,y)$ is the expansion factor and the dilated convolution kernel, respectively. The two equations, namely, (1) and (2) suggest that the dilated convolution is nothing but a 0 filling of the convolution kernel. Which in fact, can be used to increase the perception field of the convolution kernel, while also conserving the original pixel information which in turn increases the resolution? If j represents the size of the convolution kernel and k is the expansion rate, then the actual effective size of the dilated convolution becomes $j+(j-1)\times(k-1)$. Which, when compared with ordinary convolution of the same size clearly yields that expansive convolution not only helps in expanding the field of perception but also keeps the resolution same as that of ordinary convolution.

(2) Attention Mechanism: The sub-sampling process in this paper causes loss of details. To mitigate this problem, an attention mechanism is adopted, which also guides model training better. The weighted processing of feature map [8] used in the attention model can enhance the target features while suppressing the background. The term “target features” mainly specifies the contour and texture information of the eyes, eyebrows etc. from the cropped non-masked region of the face. The attention mechanism used in this paper is mainly composed of three essential parts. As shown in figure 3, namely: Spatial attention mechanism and Channel attention mechanism.

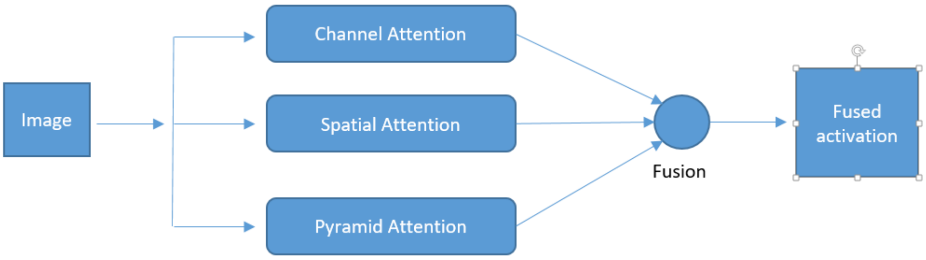

Fig. 3: Parts of the attention mechanism

(3) Spatial Attention Mechanism: The idea behind spatial attention mechanism is mainly imitating the human visual mechanism. When the human vision system looks at something, it pays more attention to some part than the other parts. When we see a bird, for example, we mainly focus on its wings as it is a key feature for recognizing a bird. Just like that in this algorithm also some parts of the feature map should have more weight that the other parts. The spatial attention mechanism proposed in this paper is presented in Figure 4.

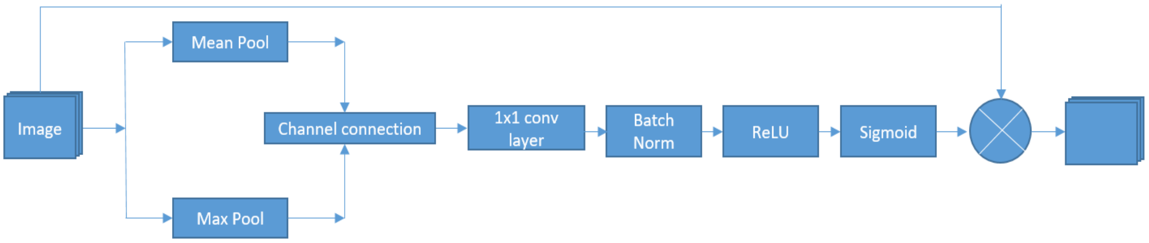

Fig 4: Spatial attention mechanism

The output featured graph uses maximum and mean global pooling on the channel dimensions. The feature weight of each position is obtained from feature information of different positions on the feature map which is also compared and extracted.By means of channel connection the two feature maps are connected in channel dimension. In order to integrate the information extracted by the two methods, the size of the 1 × 1 convolution kernel is used for learning. Feature map was calculated by sigmoid function to get the final weight of attention, so as to avoid errors caused by excessive weight coefficient. The sigmoid function is expressed as follows: where SF (x) is the response of the output and x is the input.

(4) Channel Attention Mechanism. Each and every channel of the feature map extracted by convolution neural network (CNN) represents an image feature, such as texture and shape. Here the target features are the contour and texture information of the eyes, eyebrows, and face due to the features of the face of the mask. In the image, each feature contains different information, and its contribution to image segmentation is also different. Therefore, the different attention should be measured to each different feature and different weights should be assigned. Channel attention mechanism is programmed to assign weight to features so that the network can focus on important features, as shown in Figure 6.

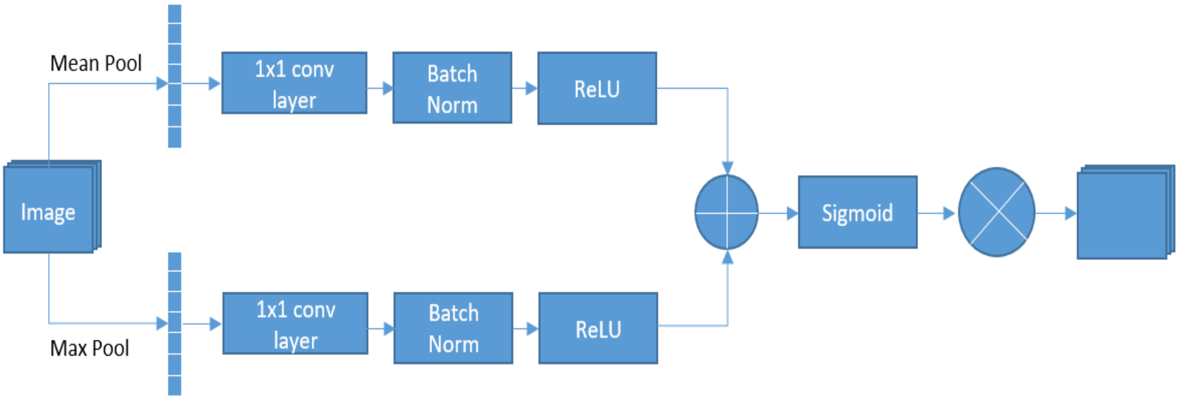

Fig 6: Channel attention mechanism.

The feature graphs obtained in order to use less computation to integrate by global pooling, the two feature graphs are, respectively, passed through a convolution kernel of size 1 × 1. Next, nonlinear components are added through the BN layer and ReLU layer to make the model fit better. This method can prevent the occurrence of over fitting phenomenon to a certain extent. Through the sigmoid function finally, the two feature graphs and weights are obtained. In the field of face detection, in general, the real deal is that faces can be in various orientations, different lighting conditions, noise, lower resolution and of course facial occlusions. Although most of these problems can be solved by selecting an appropriate position of the hardware or sensor (mostly camera) or even modern deep learning algorithms, the problem of partial occlusion still stands as a the Great Wall. It is so because there is no accurate way to find out what is behind those occlusions. In the recent years, the pandemic situation caused by COVID-19 has resulted in wearing masks to a part of our daily life. Most of the today's state of the art face detection algorithms depend on the facial features of the mouth and the nose regions (the nose region is used to correct for different projections). In a recent report of the NIST agency has shownthat many established commercial face recognition systems are showing an error of 5% – 50%. This is because these systems were designed to work on full faces (faces without mask) and now wearing masks is causing these algorithms to lose about half of the useful features, resulting in poor performances. Another major challenge is that there is not enough labeled data of masked faces, so it becomes hard to train big and complex neural networks. Although some tools have been developed to created masked faces but the distribution of these fake images and the real ones still has a questionable distance.

Application of Mask and No mask face detection:

In the current scenario everybody is wearing face masks to protect themselves from corona virus. For that the face reorganization technique is becoming weaker day by day. But by this method of masked face reorganization we can detect faces through masks. Here are some fields where this method can be useful.

1. Health Sector: In this pandemic situation people often have to visit hospitals and nursing homes. By this technology the authority can detect faces of their patients and visitors without removing the masks. That can reduce the spread of the deadly virus.
2. Government office and Private Sectors: There are certain government sectors and private farms where the authority uses face recognized check in or attendance methods. In that case the employees don't have to remove their masks in order to put their entry in the office register.
3. Security Purposes: In this mask era crime rate can go high by hiding faces behind face masks. In this case this method can be used in the security systems like road cameras and CCTV cameras to detect faces of the allegiants and this can reduce the crime rates.

4. Smart devices and Smart phones: In today's world smart devices and mobile phones are one of the most important elements. Most of the smart devices have facial recognition unlock systems. But those backdated technologies are not capable of detecting faces through masks. If this masked face reorganization is used in those smart devices, users don't have to put off their masks to unlock the devices when they are outside. This can be very useful in emergency situations.

Conclusion

In this COVID-19 situation, it has become mandatory for us to wear a mask, which makes face recognition a very challenging task. Therefore, our present face recognition methods are not able to easily recognize the face correctly. The method proposed in this paper generalizes the ways of face recognition, in so that wearing and mask does not hinder the identity. The proposed method can also be used in richer applications such as violent video retrieval and video surveillance. The proposed method achieved a high rate of recognition. To the best of our knowledge, this is the first project that addresses the problem of recognizing both masked and no-mask face recognition. It is worth it to state that this study is not limited to this pandemic period since a huge number of people are self-aware constantly, they are taking care of their health and wearing masks to protect them against pollution and to reduce other pathogens transmission. As per future scope, the subsequent works will be focused on including more occlusions and not just mask, making the model more efficient in terms of CPU and memory consumption, reducing the stages of the model, a better method for eliminating the masked region of the face.

References

[1]. King, Davis E. (2009). Dlib-ml: A Machine Learning Toolkit**.** *Journal of Machine Learning Research, 10(*06), 1755−1758.

[2]Hariri, W. (2021). Efficient masked face recognition method during the covid-19 pandemic. *Signal, image and video processing*, 1-8.

[3]. Wu, GuiLing. (2021).Masked Face Recognition Algorithm for a Contactless Distribution Cabinet. *Mathematical Problems in Engineering 2021* .

[4]. Schroff, Florian, Dmitry Kalenichenko, and James Philbin.(2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.*

[5]. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[6]. A. Vaswani, N. Shazeer, N. Parmar et al.(2017). Attention is all you need. *Proceedings of Advances in Neural Information Processing Systems*(pp. 5998–6008). Long Beach, CA, USA.

[7].davidsandberg. (2018, April 16). GitHub - davidsandberg/facenet: Face recognition using Tensorflow. Retrieved from https://github.com/davidsandberg/facenet

[8]A. Gilra and W. Gerstner. (2018).Non-linear motor control by local learning in spiking neural networks. *Proceedings of the International Conference on Machine Learning*, Stockholm, Sweden.

[9]K. He, X. Zhang, S. Ren, and J. Sun (2016).Deep residual learning for image recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). IEEE, Las Vegas, NV, USA.